

# NERSC Status and Update

Bill Kramer  
NERSC General Manager  
[kramer@nersc.gov](mailto:kramer@nersc.gov)

NERSC User Group Meeting  
September 17, 2007





# NERSC: A DOE Facility for the Future of Science



## NERSC is the #7 priority

“... NERSC ... will ... deploy a capability designed to meet the needs of an integrated science environment combining experiment, simulation, and theory by facilitating access to computing and data resources, as well as to large DOE experimental instruments. NERSC will concentrate its resources on supporting scientific challenge teams, with the goal of bridging the software gap between currently achievable and peak performance on the new terascale platforms.”

(page 21)

## NERSC is part of the # 2 priority - Ultra Scale Scientific Computing Capability

“...The USSCC, located at multiple sites, will increase by a factor of 100 the computing capability available to support open ... scientific research—reducing from years to days the time required to simulate complex systems, such as the chemistry of a combustion engine, or weather and climate—and providing much finer resolution. ...”

(page 15)



## Overall

- **A number of improvements you will hear more about**
  - **NERSC 5**
  - **High Quality Services and Systems**
  - **New Staff**
  - **New Projects**



# Number of Awarded Projects

(status at year end)

Allocation Year	Users in Production Projects	Production	INCITE & Big Splash	Startup	of which are SciDAC
2007 (as of August)	2,855	298	7	48	46
2006	2,979	286	3	70	36
2005	2,715	277	3	60	31
2004	2,654	257	3	83	29
2003	2,380	235	3	76	21





## Changing Science of INCITE

Year	Chemistry	Astrophysics	CFD	Biology	Accelerator Physics	Combustion	Climate	Fusion Energy
2004	X	X	X					
2005		X		X		X		
2006	X	X			X			
2007	X	X	X				X	X



# Changing Algorithms of INCITE

Year	Multi Physics/Multi Scale	Dense LA	Sparse LA	Spectral Methods	N-Body Methods	Structured Grids	Unstructured Grids	Map Reduce	Data Intensive
2004	X	X		X	X	X		X	X
2005	X	X		X	X	X			X
2006	X	X			X	X			X
2007	X	X				X	X		

Phil Colella's Seven Dwarfs analogy

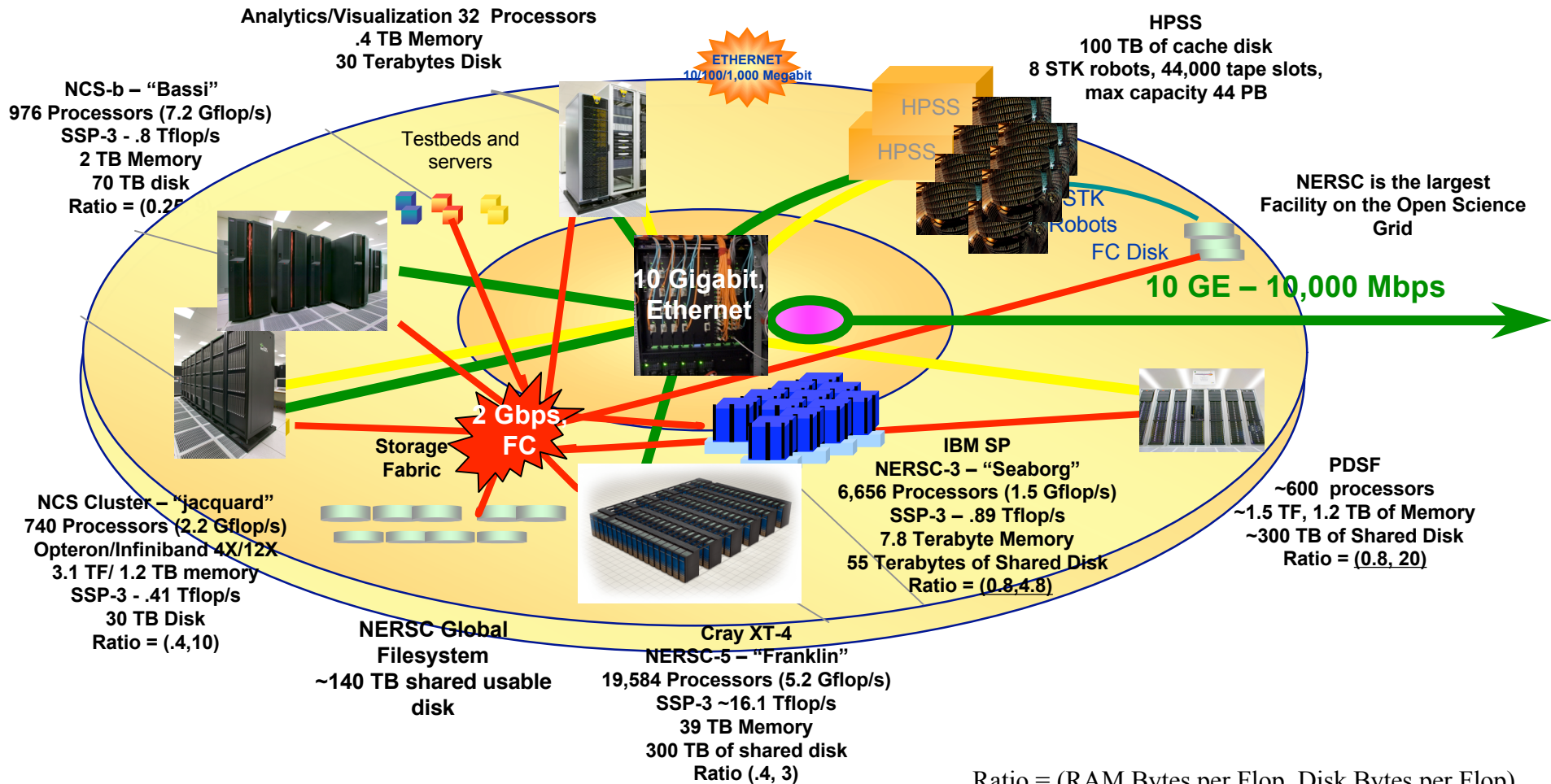


# 2007 Incite Projects

Investigator	Discipline	Location	Science Area
Larry Pratt	Chemical Science	Fisk University	Lithium Chemistry
Paul Bemis	Computer Science	Fluent Inc.	CFD for Automobile Industry
Ren Chuang	Fusion Energy	University of Rochester	Fast Ignition Simulations
Warren Mori	Fusion Energy	University of California, Los Angeles	Plasma based accelerators
Hong Im	Chemical Sciences	University of Michigan	Turbulent flame evolution
Don Lamb	Astrophysics	University of Chicago	Detonation mechanism of supernovae
Gil Campo	Climate	University of Colorado, Boulder	Global tropospheric circulation maps



# 2007






# The Real Result of NERSC's Science-Driven Strategy

## SCIENCE-DRIVEN SYSTEMS

Each year on their allocation renewal form, PIs indicate how many refereed publications their project had in the previous 12 months.



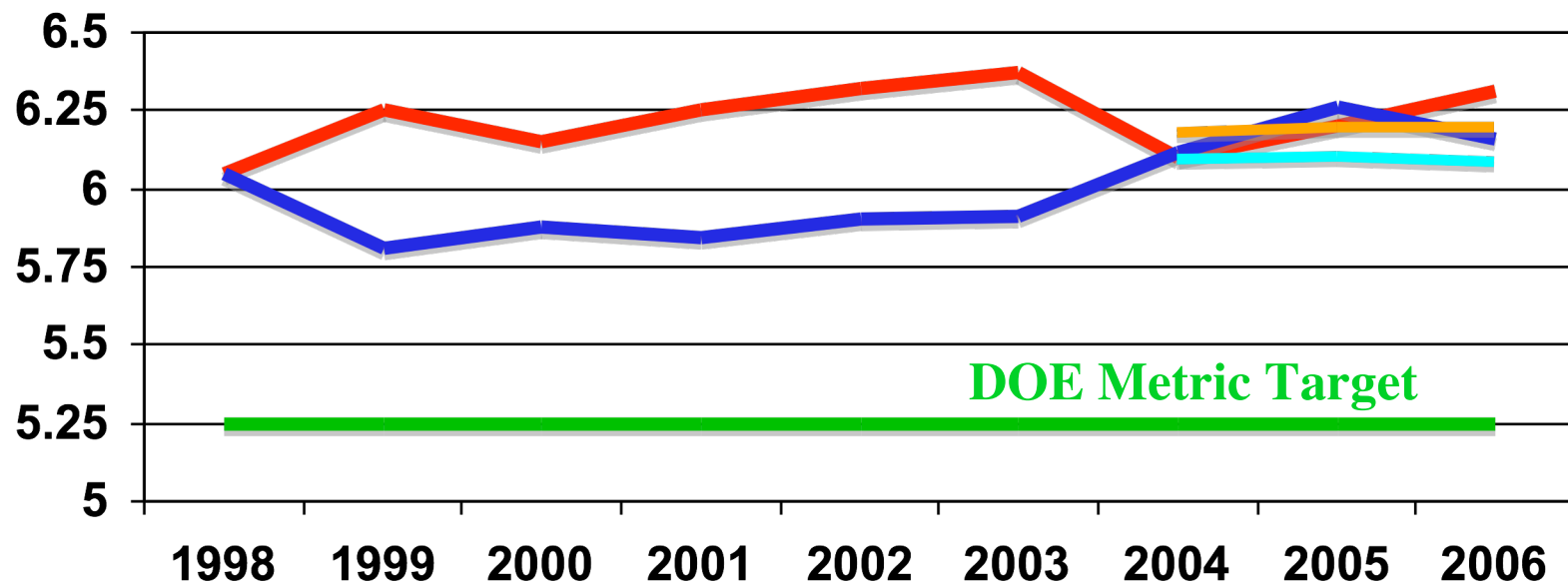
Year of request renewal submitted	Number of refereed publications
2006	1,437
2005	1,448
2004	1,270
2003	936



# User Survey Results

scores: 1 = very dissatisfied to 7 = very satisfied

- Metric 1.1.1: NERSC Overall Satisfaction - all users
- Metric 1.1.2: Average of ~10 Questions - all users
- Metric 1.1.3: NERSC Overall Satisfaction - new users
- Metric 1.1.3: Average of ~10 Questions - new users
- 5.25 - DOE Metric Target



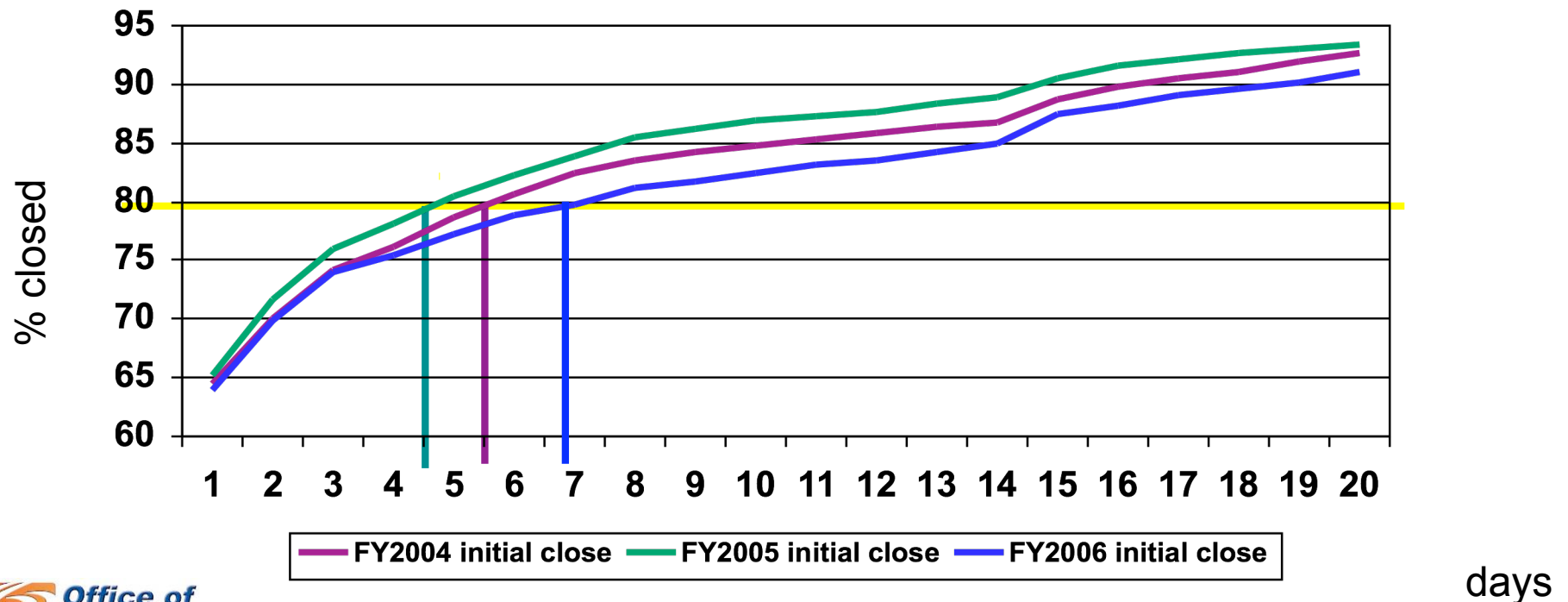




## Response Time for Assistance

Control Metric 1.2.1: 80% of user problems are addressed within 3 working days, either by resolving them or providing the user with a plan for resolution

We are implementing procedures to measure the above metric; meanwhile we show days to closure which often is significantly longer than days to a plan for resolution.





# SciDAC Collaborations

- NERSC not only supports the many SciDAC projects using its services, but participates in SciDAC projects directly that are separately funded.
- Direct Involvement
  - SciDAC Outreach Center (PI – David Skinner)
  - Open Science Grid (co-PIs Bill Kramer; Jeff Porter)
  - Petascale Data Storage Institute – (co-PIs Bill Kramer; Jason Hick, Akbar Mokhtarani)
  - Visualization and Analytics CET (co-PI – Wes Bethel)
- Close Collaborations with other SciDAC Projects
  - Science Data Management (Kurt Stockinger)
  - Performance Engineering Institute (PERI) – (David Bailey, Daniel Gunter, Katherine Yelick)
  - Advancing Science via Applied Mathematics (Phil Colella)
  - Scalable Systems Software (Paul Hargrove)



# Systems Availability/Reliability

Metric 2.1.1: Scheduled availability is  $\geq 90\%$  for systems in their first year of production,  $\geq 93\%$  in their second year,  $\geq 95\%$  thereafter.  
Metric 2.1.2: Overall availability tracked.

System	FY04 Sched Avail	FY04 Overall Avail	FY04 MTBI d:h	FY05 Sched Avail	FY05 Overall Avail	FY05 MTBI d:h	FY06 Sched Avail	FY06 Overall Avail	FY06 MTBI d:h
Seaborg	97.57	96.68	17:15	98.53	97.27	10:18	98.71	98.30	16:02
Bassi							97.0	95.64	12:01
Jacquard							99.04	97.97	8:16
Analytics Servers	99.0	98.94	21:01	99.99	99.99	29:17	99.07	98.35	11:20
HPSS user sys	99.47	98.90	8:10	99.27	98.09	8:12	99.86	98.21	6:04
NGF							99.37	99.24	23:15

MTBI - uses overall measure not just scheduled



## Systems Availability/Reliability Metrics for FY06

Control Metric 2.1.1: Scheduled availability is  $\geq 95\%$

<b>Systems</b>	<b>Scheduled Availability</b>	<b>Overall Availability</b>	<b>Mean Time Between Interrupts (Day:Hr:Min)</b>	<b>Mean Time to Restoration (Hours)</b>
MPP	98.70%	98.30%	16:02:03	7.1
Storage	99.85%	98.71%	6:08:33	2.2
Analytics	99.07%	98.35%	11:19:41	5.1



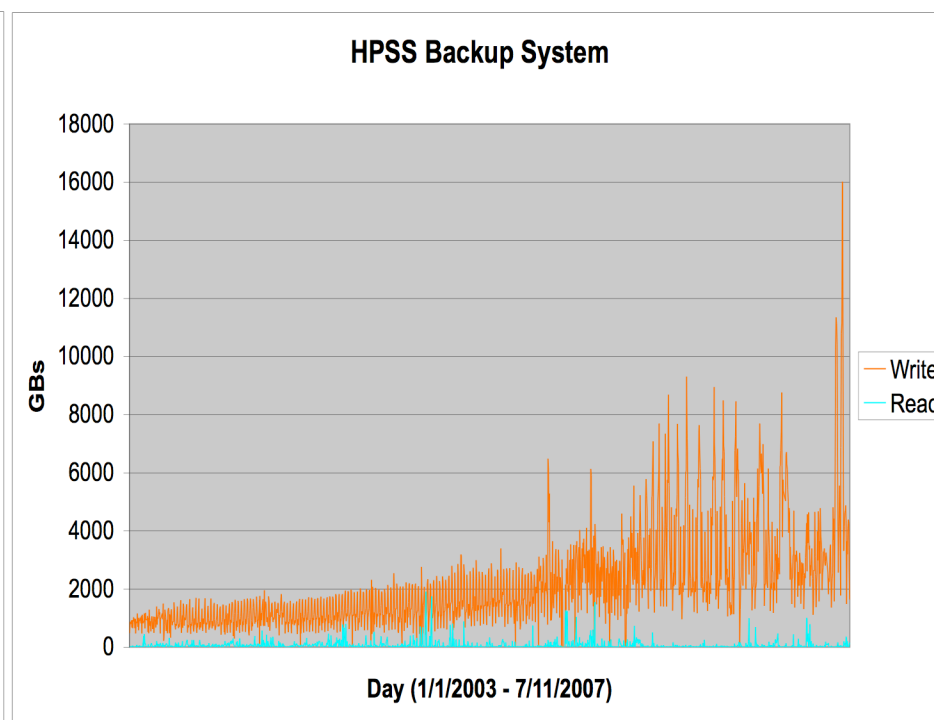
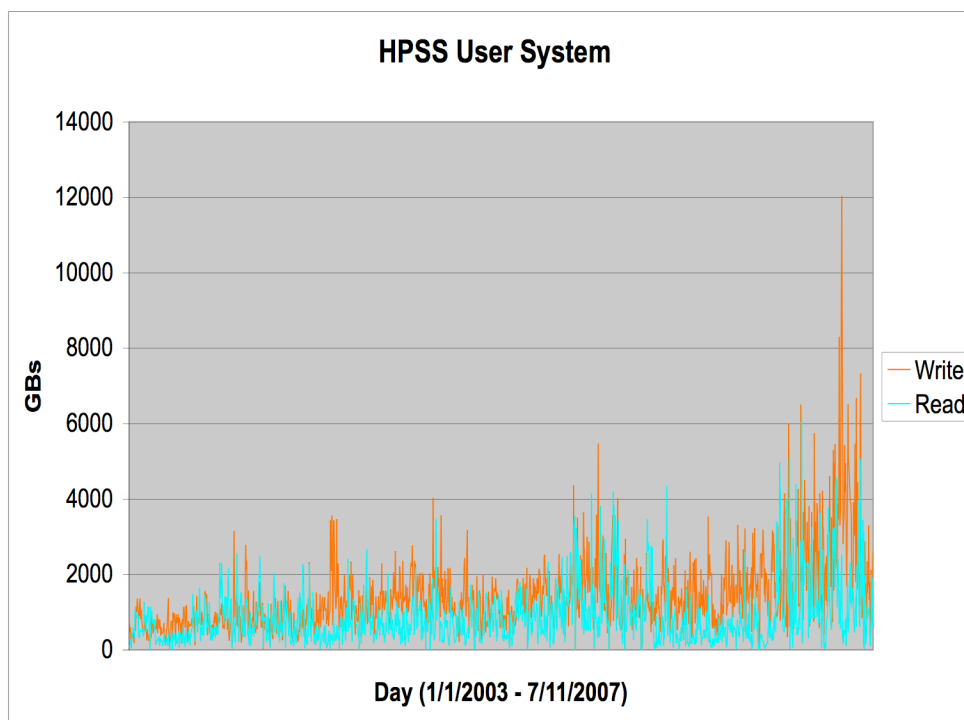
## 2006 MPP Utilization

System	Total MPP hours used	% total utilization	Total Science hours used	% science utilization
Seaborg (POWER3)	51,904,424	88.0	51,552,095	87.4
Bassi (POWER5)	42,936,702	83.1	42,107,990	81.5
Jacquard (Linux cluster)	19,846,287	79.9	19,495,570	78.4

Duty Cycle Target is 80-85%



# Daily HPSS I/O



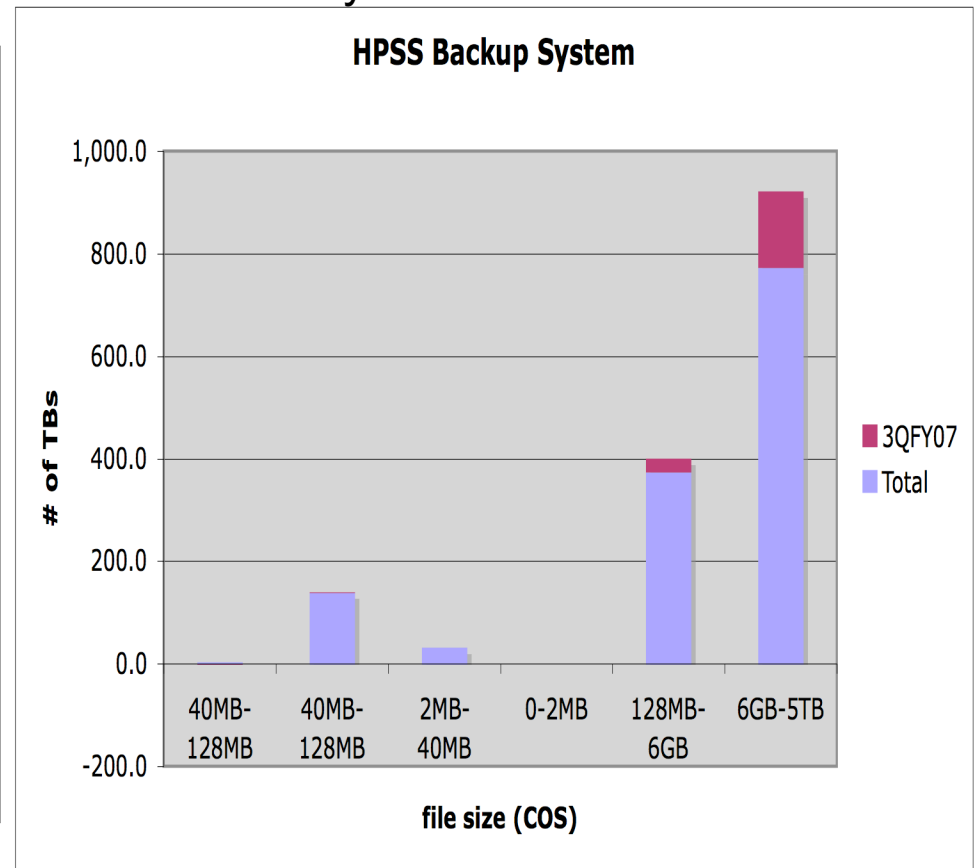
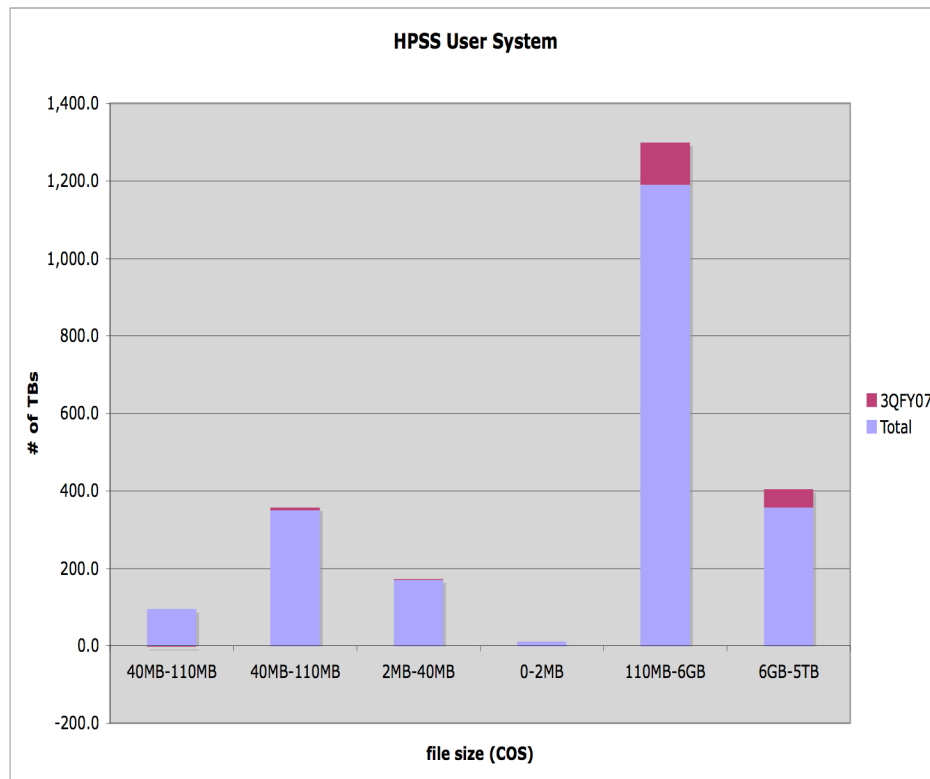




# HPSS Data Distribution

- User system (1/1/2007-8/20/2007)
  - 3,730,710 new files
  - 447 terabytes of new data

- Backup system (1/1/2007-8/20/2007)
  - 847,228 new files
  - 307 terabytes of new data

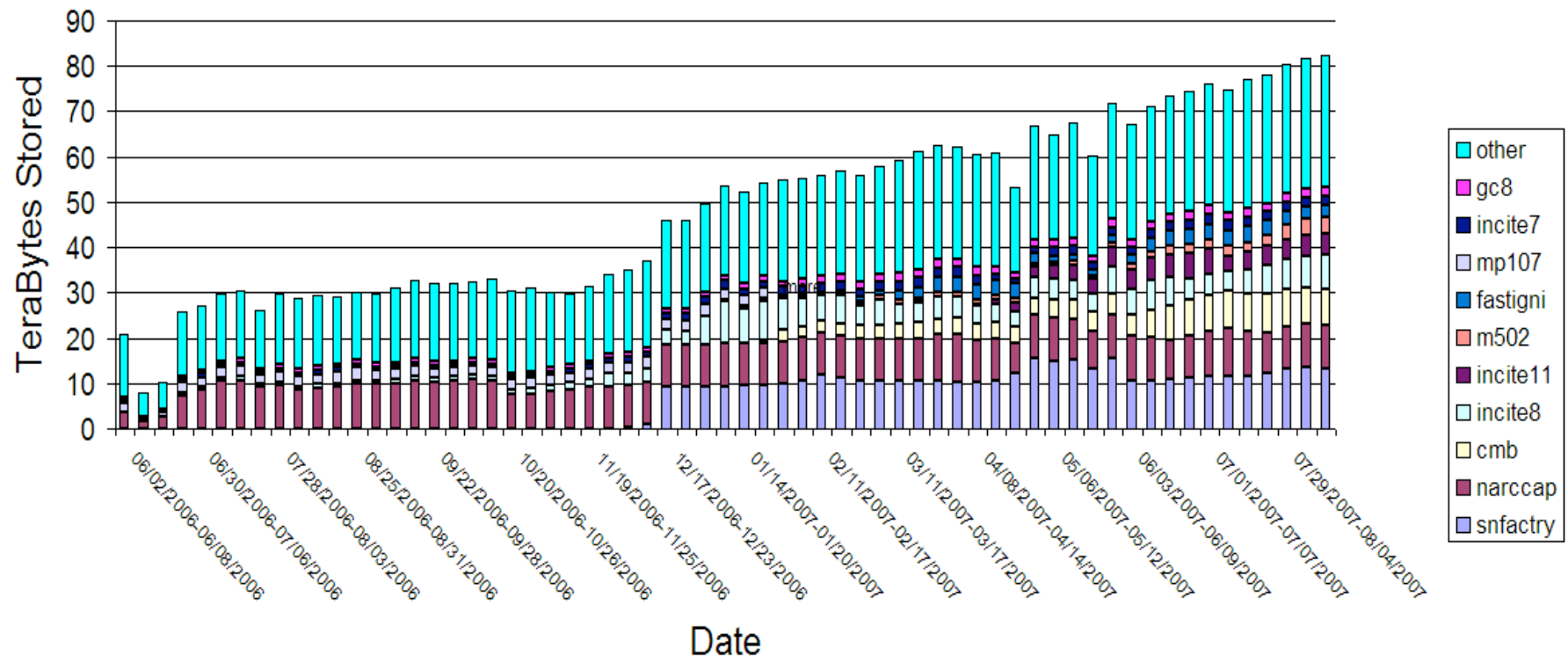




# NERSC Global Filesystem (NGF) Utilization Collection

NGF staff collect the amount of data stored and number of files per project in NGF. There are 85 projects using NGF.

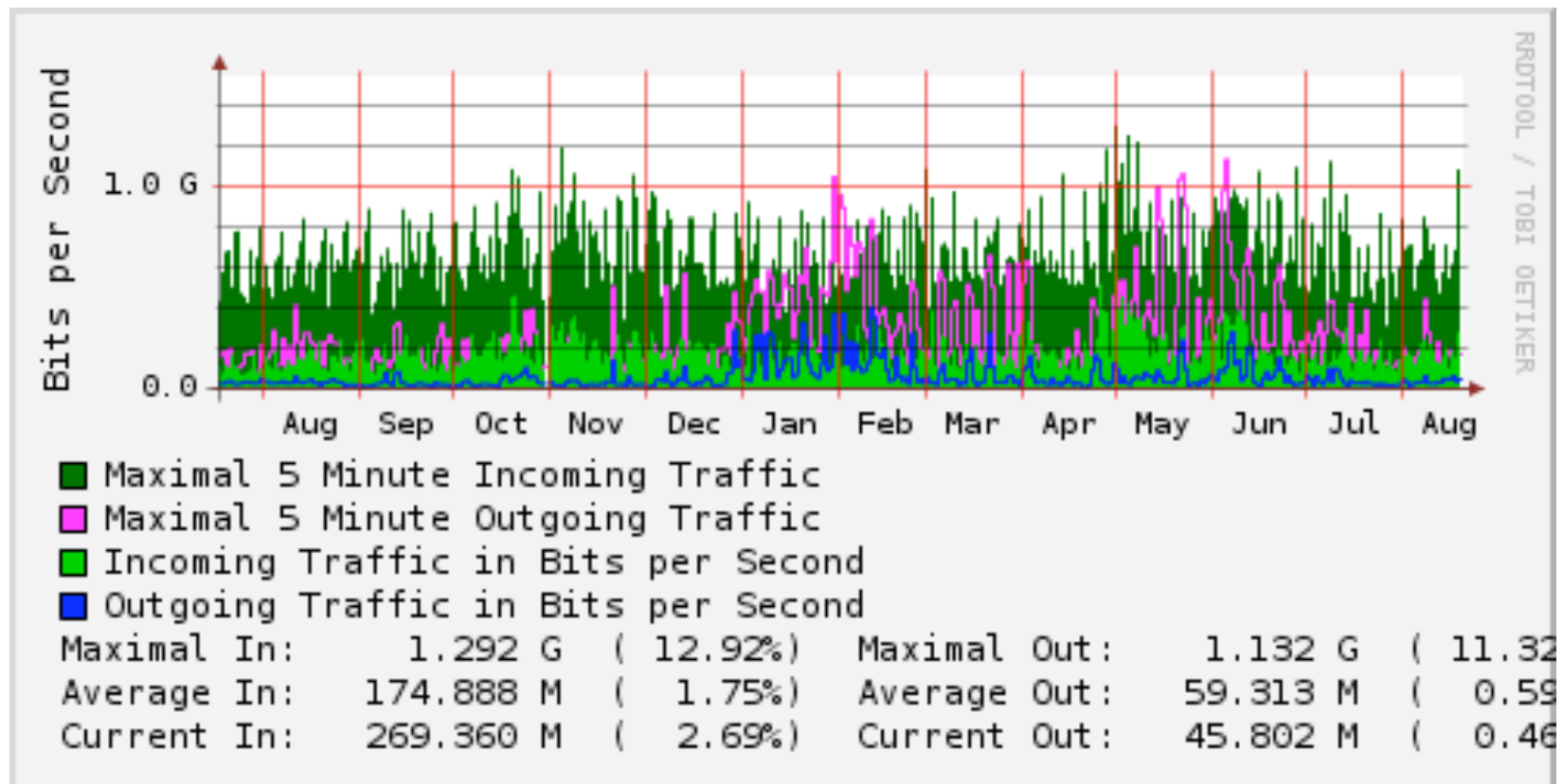
NGF space usage(TB)





# Network Resource Utilization Collection

Networking staff collect data on amounts, rates, and errors coming in/out of NERSC and from internal networks.

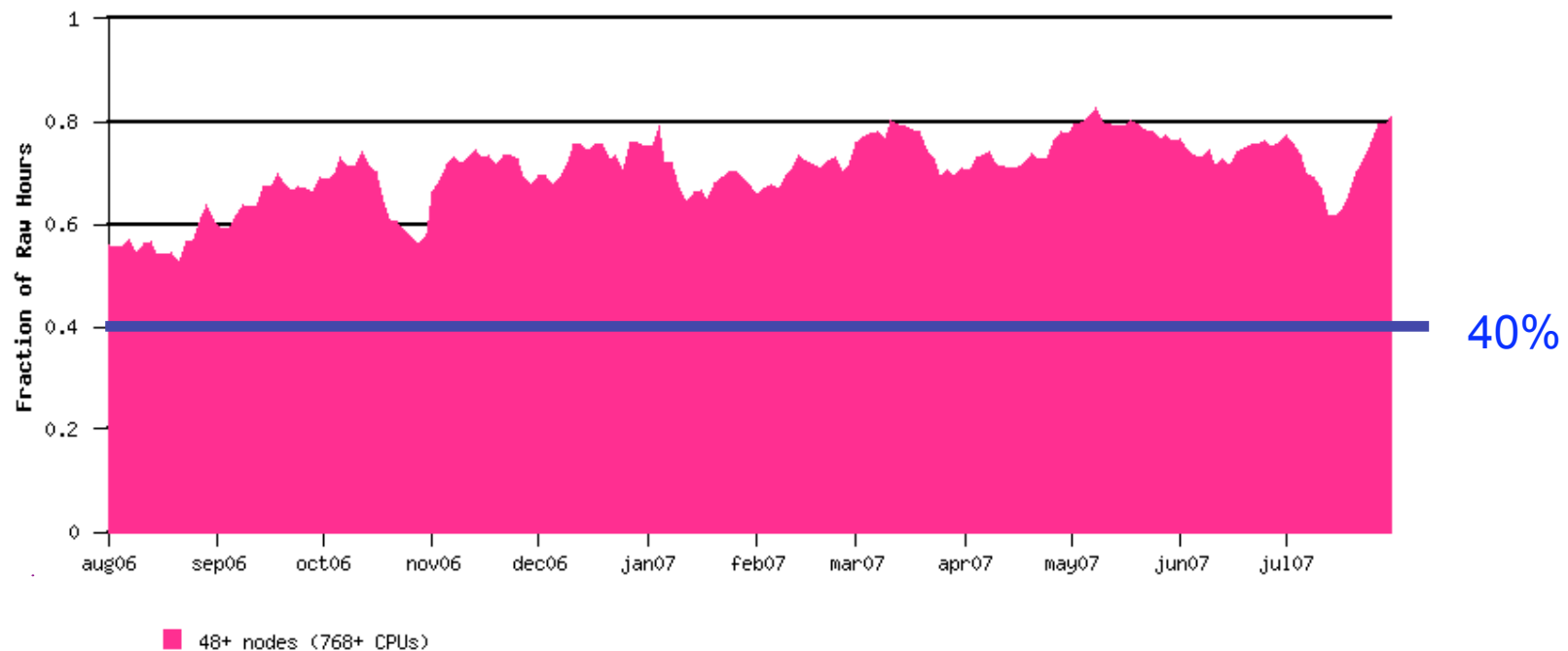




## Priority Service to Capability Users

Control Metric 2.3.1: on capability machines at least 40% of the cycles should be used by jobs running on 1/8<sup>th</sup> or more of the processors.

The graph shows the percent of Seaborg cycles run on 1/8<sup>th</sup> or more of the processors. About half of these “big cycles” were provided by the DOE allocation; half by incentive programs.





## Job Throughput of Capability Jobs

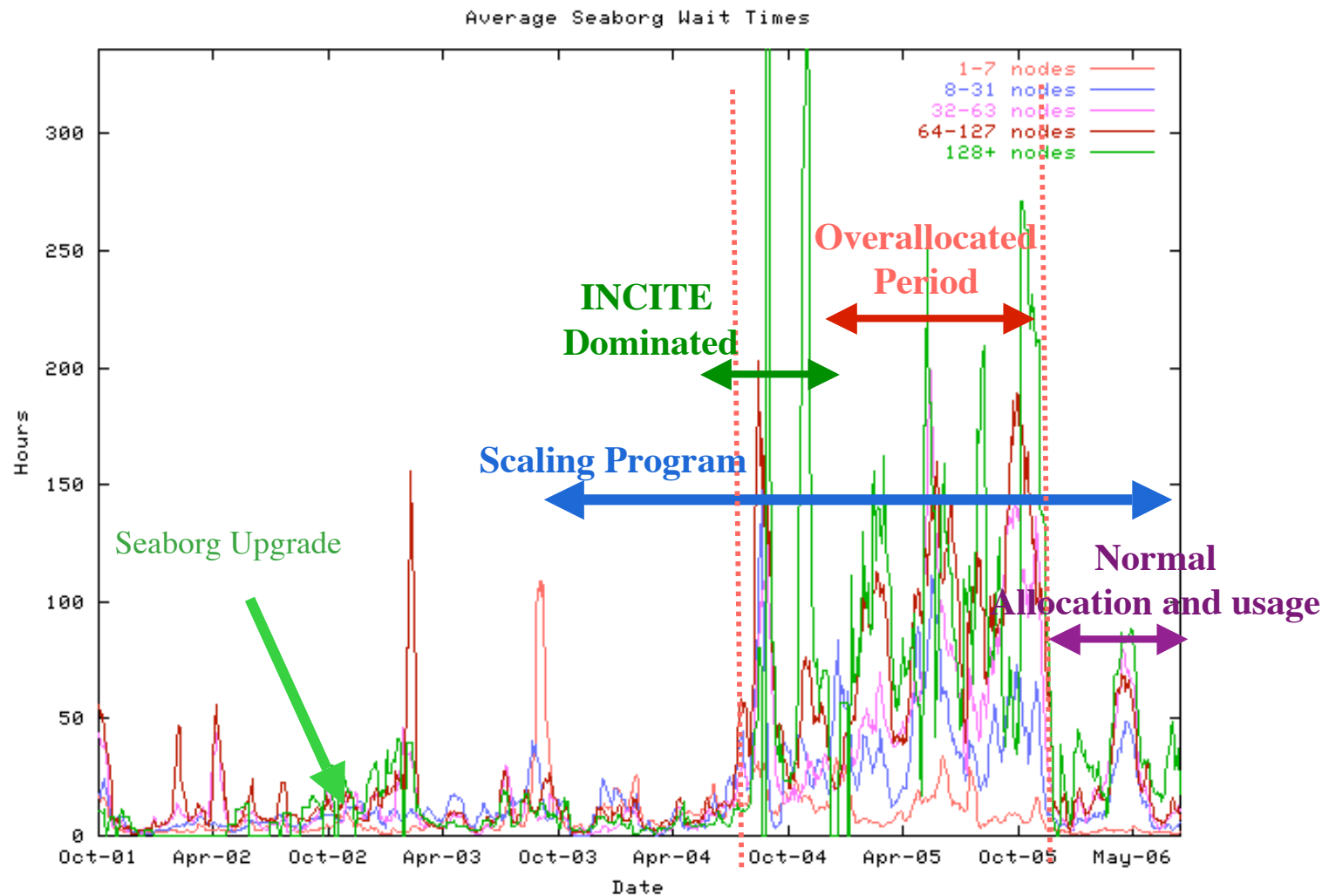
### Control Metric 2.3: NERSC tracks job throughput

The table below shows the expansion factor (EF) for Seaborg's regular priority capability jobs.  $EF = (\text{wait time} + \text{requested time}) / \text{requested time}$

	Allocation pressure	512-1008 procs EF	1024 – 2032 procs EF	2048+ procs EF
FY04	INCITE pressure	4.67	7.21	6.98
FY05	Significant over-allocation	6.51	8.84	13.68
FY06	Average	3.58	4.13	4.96
FY07	Pent up demand?	5.25	5.59	9.85
Aug 1-22	Big users --> Franklin	2.27	2.47 (avg wall: 12 h 10)	8.78 (avg wall: 3h 26)



## 4 Year Seaborg Queue Wait Statistics

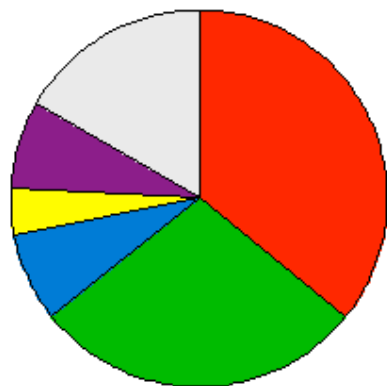






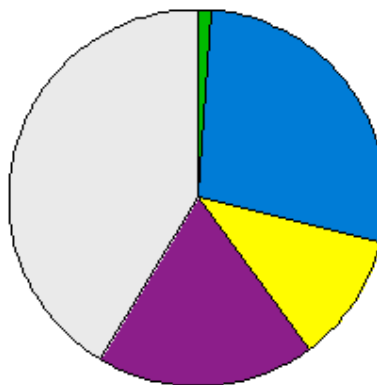
# FY2007 (thru Aug 20) Usage by Job Size by System

Seaborg -  
380 nodes  
6,080 procs



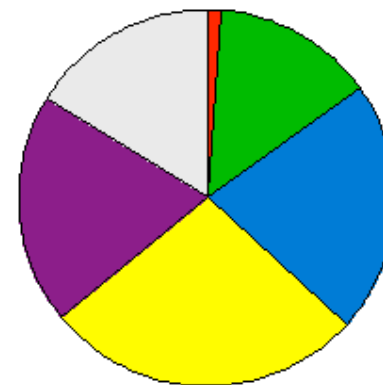
- 128+
- 64-12
- 32-63
- 16-31
- 8-15
- 1-7 N

Bassi -  
111 nodes  
888 procs

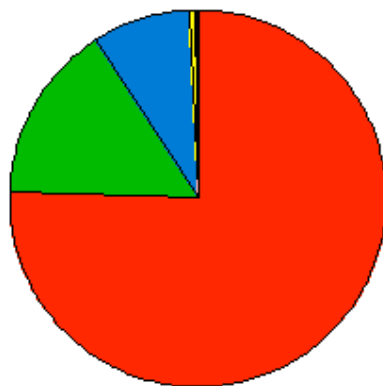


- 64+ Nodes
- 32-63 Nodes
- 16-31 Nodes
- 8-15 Nodes
- 1-7 Nodes

Jacquard -  
356 nodes  
712 procs



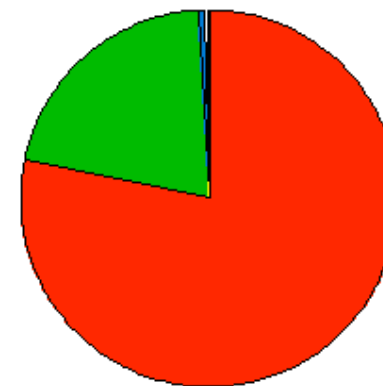
- 128+ Nodes
- 64-127 Nodes
- 32-63 Nodes
- 16-31 Nodes
- 8-15 Nodes
- 1-7 Nodes



- DOE Ba
- SciDAC
- INCITE
- Overhe
- Startu
- SC Dir
- Guest



- DOE Base
- SciDAC
- INCITE
- SC Direct
- Overhead
- Startup
- Guest



- DOE Base
- SciDAC
- Overhead
- Startup
- Guest
- INCITE



## Projects are Sharing Data Sets

Project	NERSC Users	Data (TB)	Uses of the Data
STAR simulations and detector data analysis	297	540	study the collision of heavy nuclei
Kamland simulations and detector data analysis	50	438	measure neutrino oscillation parameters
Applied Partial Differential Equations Center	20	357	develop high-resolution adaptive methods for PDEs applied to combustion and astrophysics applications
Program for Climate Model Diagnosis (PCMDI)	119	219	Half is raw data and half is a mirror for IPCC data served to the public by the Earth System Grid. Future plans to serve data directly from HPSS.



## Projects are Sharing Data Sets

Project	NERSC Users	Data (TB)	Uses of the Data
Supernova Factory	31	173	discover supernovae, measure the expansion of the universe and dark energy
Precision Cosmology (2006 INCITE)	4	100	study cosmic dark energy
Joint Genome Institute	20	84	repository of gene sequence data, may be served to public in the future
Supernovae Detonations (2007 INCITE)	14	76	study white dwarf stars
DNA Damage and Repair	3	66	Nuclear Imaging of Gene Expression

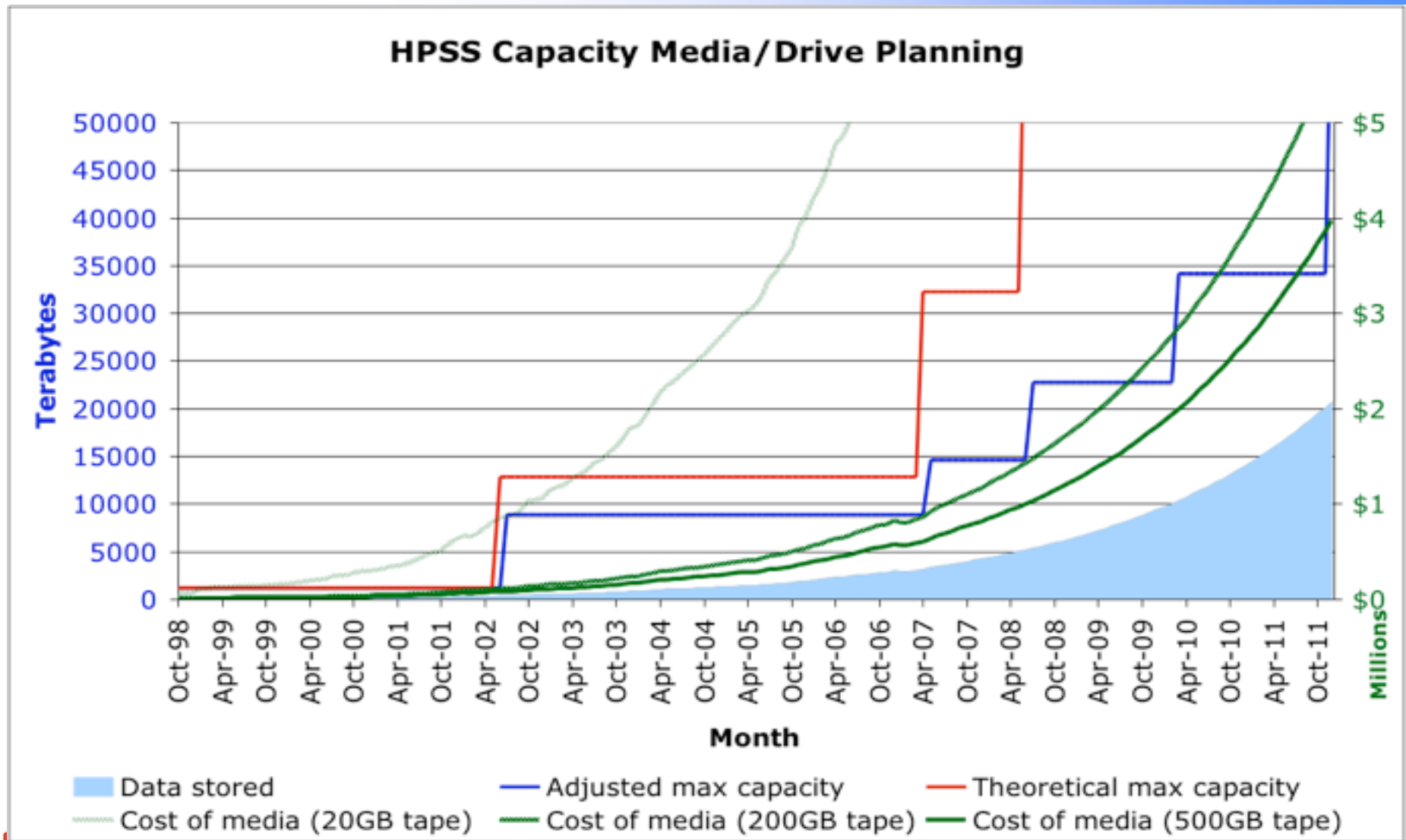


## Projects are Sharing Data Sets

Project	NERSC Users	Data (TB)	Uses of the Data
Deep-Sky map	-	60	observations taken by Palomar-QUEST Consortium, SN Factory, Near Earth Asteroid Team; plans to serve to public
Cosmic Microwave Background Data Analysis	116	37	answer questions in cosmology, fundamental physics, and astrophysics
Molecular Dynameomics (2005 INCITE)	-	6	characterize the native state dynamics and the folding / unfolding pathway of proteins; data served to public at: <a href="http://www.dynameomics.org">www.dynameomics.org</a>
Lattice Gauge Theory	-	3	NERSC hosts a repository of gauge configurations for Columbia QCDSP and MILC: <a href="http://qcd.nersc.gov/">http://qcd.nersc.gov/</a>



# Mass Storage Strategy: Media / Drive Planning





# FY 07, FY 08 and Beyond





# FY 07 Accomplishments

## Beyond Science done at NERSC

- **Delivery, testing and deployment of the world's largest Cray XT-4**
  - Made more interesting when we switched from a CVN acceptance to an Compute Node Linux Acceptance - first site to run full time at scale
- **Site Assist Security Visits - with very good results**
- **More hours, users and projects than ever before**
- **All system meeting and exceeding goals**
- **NERSC Global Filesystem impact**
- **Scaling Program - 2**
  - We worked on the obvious areas in SP-1 - most projects have qualified for leadership/INCITE time.
  - Now we are working on the areas that are “the high hanging fruit”
- **SDSA Impact**
  - Berkeley View paper, Cell, Multi-core studies
- **Design of GPFS/HPSS interface with IBM**



# FY 07 Accomplishments

## Beyond Science done at NERSC

- **Delivery, testing and deployment of the world's largest Cray XT-4**
  - Made more interesting when we switched from a CVN acceptance to an Compute Node Linux Acceptance - first site to run full time at scale
- **Site Assist Security Visits - with very good results**
- **More hours, users and projects than ever before**
- **All system meeting and exceeding goals**
- **NERSC Global Filesystem impact**
- **Scaling Program - 2**
  - We worked on the obvious areas in SP-1 - most projects have qualified for leadership/INCITE time.
  - Now we are working on the areas that are “the high hanging fruit”
- **SDSA Impact**
  - Berkeley View paper, Cell, Multi-core studies
- **Design of GPFS/HPSS interface with IBM**



# FY 08 Plans

- **Full Production with NERSC-5**
  - Major software upgrade in June 2008
    - Checkpoint Restart
    - Petascale I/O Forwarding
    - Other CNL functionality
  - If it performs as expected will upgrade Franklin to Quad core
    - Total of 39,320 cores
- Upgrade to NGF-2 to fully include Franklin
- Deploy new analytics system (in procurement now)
- Upgrade/balance MSS and network
- Focus on scalability with users on Franklin
- Start NERSC-6 procurement - DME - for CY 2009 delivery
  - Revise NERSC SSP benchmarks
- Support new user communities
  - NOAA, NEH, others?
- Provide excellent support and service



## FY 09-13 Plans

- **Procure and Deploy NERSC-6 - initial arrival in CY 2009**
- **Move to new CRT building on site - 2010-2011**
- **Center balance**
  - Replace tape robots and keep up pace with storage
  - Upgrades to LAN and WAN along with Esnet
  - NGF expansion
  - Analytics and infrastructure
- **NERSC-7 in 2012 (first new system in CRT)**
- **Excellent support and service**



# NERSC Long Range Financial Plan

- NERSC's financial plan (FY06 to FY12) is based on DOE's budget request to OMB
  - FY07 budget was reduced from \$54,790K to \$37,486K
    - due congressional delays in passing a budget in 2007
  - Increase planned in FY08 to 54,790K, sustained to FY12
    - Necessary to meet performance goals
    - NERSC's cost plan meets budget request
- NERSC was able to absorb the reduction with little user effect by
  - Capping staff growth
  - Deferring payments on NERSC-5
  - Cutting Center Balance funding
  - Other reductions

	Actual		Projected Plan					
	FY06	FY07	FY08	FY09	FY10	FY11	FY12	Total
Budget	38.6	37.5	54.8	54.8	54.8	54.8	54.8	350.1



# Risk Management Plan

## FY08 Budget Risk

Budget impact to NERSC during Continuing Resolution in FY08, if NERSC budget remains at FY07 levels i.e. \$37.5M:

### Response:

- **Eliminate Center Balance and other improvement activities**
  - No improvements to HPSS, delay NGF or Networking activities
  - Cancel the DaVinci replacement
  - Decommission Jacquard and Bassi, saves electricity and maintenance
- **Reduce Staff**
  - There is a long term impact to services, recovery would not be immediate when budget is restored
  - Commitments to NERSC 5 and lease are firm and costly to renegotiate
- **Additional budget trimming required ~500K, most likely will delay activities related to the next power upgrade**
- **Impact to DOE OMB goals:**
  - Allocation hours decrease from 450M CRHs to 405 CRHs in FY08,
  - Remains at 405M CRHs through FY09
  - Reduce OMB goal 1,200M CRHs to 725M CRHs in 2010





## Aerial View



LBNL-CRT

PERKINS  
+ WILL



# Summary

**We have made a lot of improvements this year**

**You will hear about exciting things for the rest of today**

**NERSC values the feedback you will give us today and always**

**We have worked together to facilitate and produce new science**

**We need your help to keep NERSC strong and vital**